

A SEMI-AUTOMATIC TOOL FOR META-DATA EXTRACTION FROM MALAWI COURT JUDGMENTS - 1ST AFRICANLP WORKSHOP, INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR 2020)

Amelia Taylor;

;

© 2021, AMELIA TAYLOR



This work is licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly credited.

Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/legalcode>), qui permet l'utilisation, la distribution et la reproduction sans restriction, pourvu que le mérite de la création originale soit adéquatement reconnu.

A Semi-Automatic Tool for Meta-Data Extraction from Malawi Court Judgments

AI4D Funded Project

14th April 2020

Presenter: Amelia Taylor

Project Scope

DATA:

- Malawi Criminal Court Cases over the period 2010 -2019 (at the moment we have 665)
- Some cases are available on the platform MalawiLii
- Most cases are not available online but are held by the High Court Library in printed format
- Scanned copies of these judgments are made available by the Library to subscribers

LEGAL RESEARCH IN Malawi:

- 1) **Malawi Legal System is based on Case Law**
- 2) The Laws of Malawi – are available online but not maintained regularly
- 3) **Malawi has a very limited number of legal commentaries** (commercial laws).
None are available for Criminal Law.
- 4) **The classification of the judgments is rudimentary** and is done by information held in the title of the file: type of case (criminal, civil, commercial), issuing court , year and some of the parties (appellants and respondents).

Legal Research

- A legal practitioner typically wants to find relevant cases, laws and statutes.
- These can be specific statutes on a subject or a general overview of the law and how it is applicable.
- The results of a search depends on the suitability of the search keywords, e.g., how frequent they are, whether they have more than one meaning, whether they are legal terms and how applicable they are to the subject matter.
- The combination of indexes, thesaurus of legal terms, an organised system of citation/ referencing is crucial to legal research.

Problem Statement

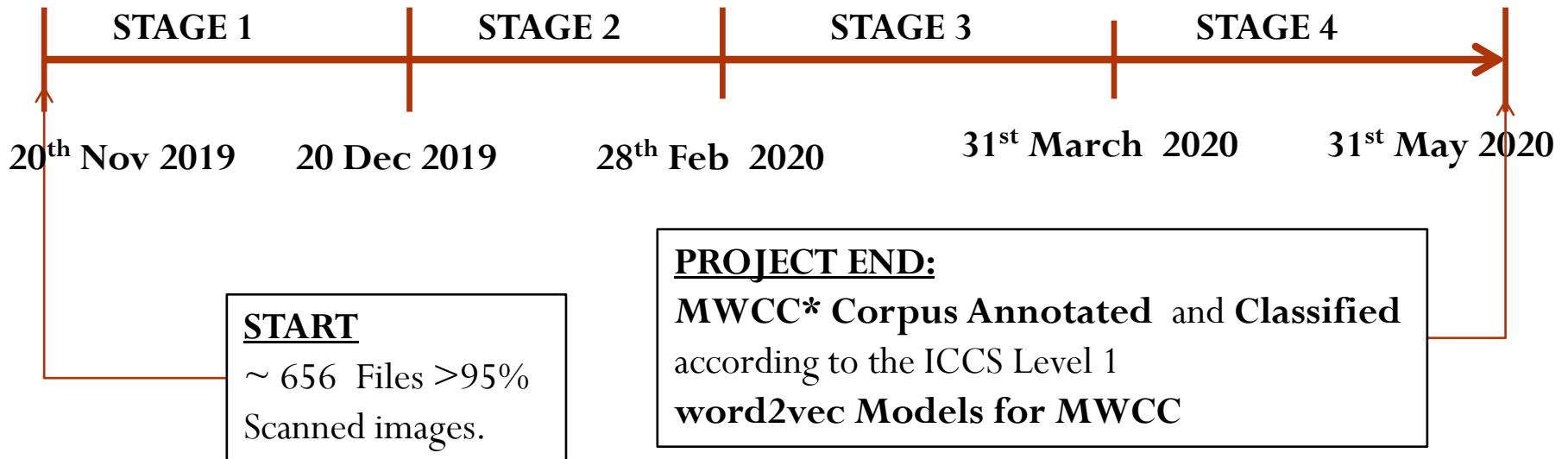
- MalawiLII is not complete and not easily searchable (most judgments are images of scanned pdfs)
- There is an absence of content meta-data, keynotes and summaries which makes legal research very challenging
- There is no linkage between cases citations
- There is no linkage from cases to Laws/ Statues
- There are no commentaries or digests for Malawi Criminal Law
- Printed Malawi Law Reports has a large gap over the last 10 years

Project Aims

1. To build an automatic tool for identifying and extracting the general structure of criminal court judgments in Malawi.
2. To build a semi-automatic tool for extracting key-meta data from court judgments.
 - Case type / court of hearing / dates / judge
 - Legal Parties involved
 - Case Citations
 - Laws/Statutes Citations
 - **Key legal terms**

Project Timeline

This has been funded by Artificial Intelligence 4 Development (Ai4D) programme as part of the 1st Call for Project proposals on Artificial Intelligence 4 Development Technologies.



MWCC = Malawi Criminal Cases

Project Parties

- Amelia Taylor, lecturer the University of Malawi: algorithms for data cleaning/ OCR/ Annotations / corpus creation / classification
- Eva Mfutso-Bengo: Lawyer and lecturer in Law at the University of Malawi – manual mapping of cases/ Laws to the ICCS classification
- Binart Kachule: - works for the Malawi Judiciary, and was an MSc student under the supervision of A. Taylor (now successfully graduated): data check and manual annotations

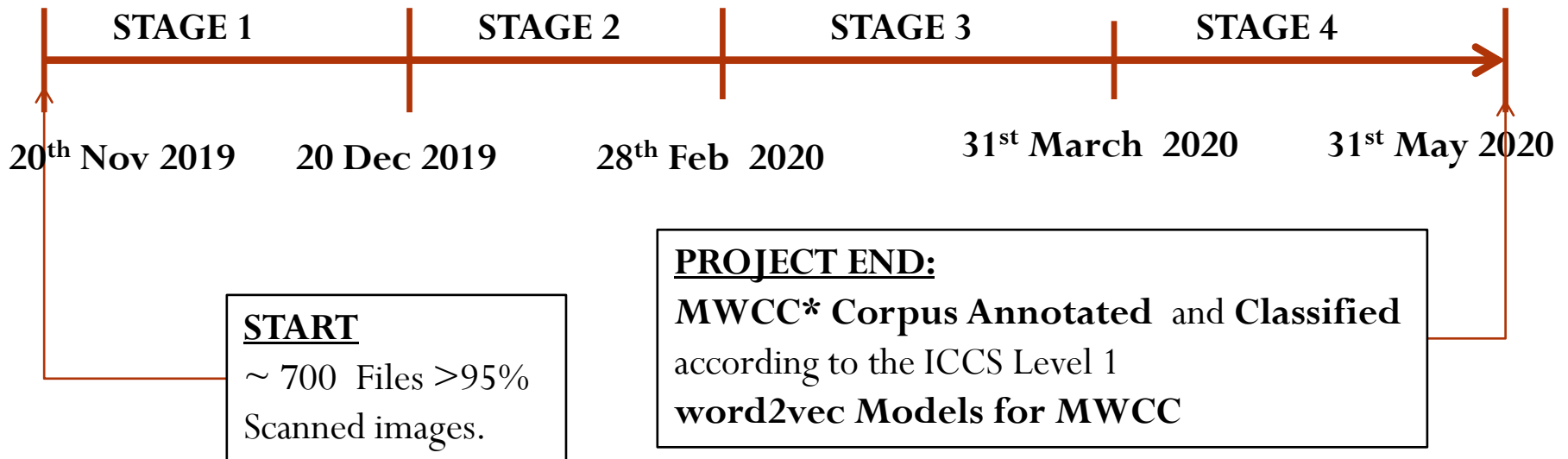
Project Timeline

STAGE 1

OCR: TEXT files for the images - line by line. Text without headers/footers/ page no / footnotes
-Watermarks, marks/stamps, orientation, marks due to dirty scanner

MWCC CORPUS - Creation of corpus files; Corpus Analysis: word frequencies/ collocations

ICCS Classification: Manual classification of 2019 cases according to ICCS Level 1, Manual mapping of the Malawi Penal Code to the ICCS Level 1



Project Timeline

MWCC CORPUS - Creation of annotated corpus files

- Which annotation standards to use: T1 Section 10(a) of Penal Code 100, 138
- Or an XML based annotation scheme, e.g., TEI

ICCS Classification: Manual classification of 2019 cases according to ICCS Level 1, Manual mapping of the Malawi Penal Code to the ICCS Level 1

- Legal Research for guiding the selection of features to be used in the training data and classification

JUDGMENT SECTIONS:

Introduction: dates, court, coram

Body: main text of the judgment

Conclusion: Incl. Date/signature

Project Timeline

TRAINING SET ANNOTATIONS:

NE: people **Chichewa names are not easily recognised by existing models.**
 NE: organisations
 NE: Money/Products - **collapsed into annotations to avoid full-stops issues.**
 NE: places
 Case Citations
 Law /Statues Citations; ICCS Legal Terms

ICCS Level 1 annotations: Selections of relevant features/ testing
 - This involves the use of collocations/ thesaurus to build a Gold Standard of legal terms

The use of Annotations to test and train models
 Create **Word2Vec Models** suitable for legal text

Experiment with various paragraphs sizes
Long judgments/ long paragraphs / in-case quotes

OCR Challenges

HC/PR Conf. 185/13
Rep v Tchasumbu 3

IN THE HIGH COURT OF MALAWI PRINCIPAL REGISTRY CONFIRMATION CASE NO. 1102 OF 2001

The Republic Vs Stance Chikopa and Ibrahim Kantande

From the First Grade Magistrate Court sitting at Dalton Road,
Limbe in Criminal Case No. 296 of 2000

CORAM: CHIMASULA PHIRI J.
Miss Nayeja – Counsel for the State
Accused – Both present/unrepresented
C. K. Chisi - Official Interpreter.

ORDER

The two accused persons were convicted after a full trial with the offence of armed robbery contrary to Section 301 of the Penal Code. Stance Chikopa was sentenced to 5 years imprisonment with hard labour while Ibrahim Kantande was sentenced to 8 years imprisonment with hard labour. The reason for this disparity is that the second accused has a previous conviction of similar type. The matter has been set down to enhance sentences.

The reviewing judge expressed his sentiments that the sentence ought to be enhanced. The State is of the similar opinion.

The first accused pleaded for mercy. He stated that he has learnt a bitter lesson in prison and is infected with TB.

Semi-Automated Lesson for Meta-Data Extraction from Malawi Court Judgments

being observed or interrupted'.³ The burglary of a dwelling house is a serious offence both by its nature and for its effect on the victims as 'it is the defilement of the house which made some householders become afraid of their own homes'; *Republic v Chizumila and others*⁴ and *Chitsonga v Republic*.⁵ The maximum penalty for committing the offences of burglary and housebreaking is death or imprisonment for life and deserve long custodial punishment: *Republic v Tomasi*.⁶ The long and immediate sentences of imprisonment are justified on the grounds that the offence of burglary is among the offences regarded as grave under our penal provisions and that burglary leaves victims and the society at abject insecurity: *Republic v Nyungwe*.⁷ In *Republic v Gobe*⁸ it was stated that 'it must be in extremely rare cases that a sentence of burglary should be suspended'.⁹ The *Magistrates' Court Sentencing Guidelines*¹⁰ suggests the starting point for the punishment of housebreaking and burglary as a sentence of six years imprisonment. The way to determine a starting point sentence for the offence has also been discussed in the cases of *Republic v Mpinganjira Bagala*¹¹ and *Republic v Dingu James Maya*.¹²

The case law provides guidance on appropriate punishments that are actually imposed in cases of similar nature. In the case of *Republic v Chizumila and others*¹³ the court suggested that the appropriate 'starting point for a down-the-mill kind of burglary'¹⁴ should be six years imprisonment. In *Republic v Kachingwe*¹⁵ it was stated that 'the starting point should be reduced or increased to reflect mitigation or aggravation'.¹⁶ Accordingly, 'for a middle of the road burglary with strong mitigating factors and no aggravation should attract a sentence of three years' imprisonment with hard labour': *Republic v Tito*.¹⁷ In *Republic v Nyungwe*¹⁸ the court outlined the crucial factors that a court should consider when sentencing for burglary to be as follows:

'... the extent of the trespass and the circumstances around the crime. These include the manner and extent of the breaking and entry. If more people were involved, obviously the court has to consider that. There could be considerable damage to the premises as entry is gained. There could be disturbance to the occupants as the crime

³ Eric Peber, Patrick Barton and Lameck Gondwe *Crimes of Need, results of the Malawi National Crime Victimization Survey* (Zomba: National Statistical Office of Malawi, 2004) at 29.

⁴ [1994] MLR 288 (HC) at 307.

⁵ [1995] 1 MLR 86 (HC) at 88.

⁶ [1997] 2 MLR 70 (HC) at 72.

⁷ [1997] 2 MLR 127 (HC) at 129.

⁸ [1995] 2 MLR 726 (HC) at 727.

⁹ [1995] 2 MLR 726 (HC) at 727.

¹⁰ Malawi Judiciary (May 2007) at 38.

¹¹ HC/PR confirmation case no. 24 of 2011 (unreported: 11 July 2013).

¹² HC/PR confirmation case no. 178 of 2013 (unreported: 21 August 2013) at 3.

¹³ [1994] MLR 288 (HC) at 307.

¹⁴ *Republic v Gobe* [1995] 2 MLR 726 (HC) at 727.

¹⁵ [1997] 2 MLR 111 (HC).

¹⁶ [1997] 2 MLR 111 (HC) at 112.

¹⁷ [1995] 2 MLR 638 (HC) at 644.


¹⁸ [1997] 2 MLR 127 (HC) at 129.

Malawi Criminal Cases Corpus

- Other Corpora

- British Law Report Corpus (BLaRC)
- House of Lords Judgments (HOLJ)
- Corpus of US Case Law (CUSC)

- MWCC: we started with 401 Judgments – later we added to 665 Cases over 2010 – 2019 year of publishing pertaining to 36 judges.

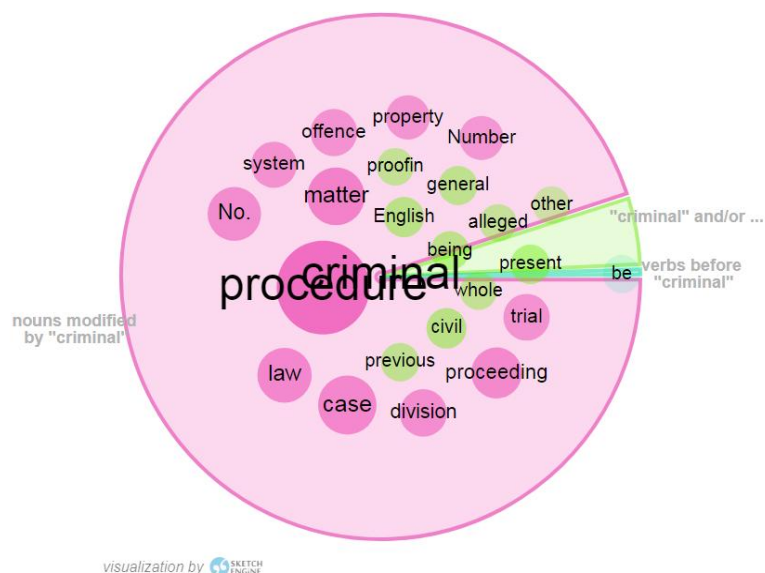


YEAR	NO_CASES
2010	85
2011	63
2012	21
2013	153
2014	85
2015	122
2016	46
2017	27
2018	42
2019	21
Total Result	665

Corpus Analysis

20 Most Frequent Verbs			Rank in terms of Freq in the whole corpus			20 Most Frequent Nouns			Rank in terms of Freq in the whole corpus			20 Most Frequent Nouns in BLRC			Rank in terms of Freq in the BLRC		
1	be	15				1	court	10				1	case	33			
2	have	33				2	appellant	22				2	court	36			
3	do	141				3	case	24				3	section	44			
4	accuse	3963				4	evidence	26				4	appeal	45			
5	say	243				5	Court	10				5	evidence	46			
6	make	208				6	person	38				6	act	48			
7	find	222				7	offence	36				7	claimant	54			
8	take	270				8	sentence	41				8	decision	58			
9	give	310				9	section	35				9	tribunal	60			
10	state	40				10	appeal	37				10	order	62			
11	go	414				11	law	50				11	appellant	70			
12	consider	283				12	fact	92				12	lord	75			
13	prove	174				13	year	693				13	law	77			
14	follow	1008				14	accused	21				14	time	80			
15	convict	129				15	Code	63				15	judge	87			
16	commit	503				16	State	40				16	article	89			
17	see	1087				17	trial	68				17	part	95			
18	tell	1243				18	Republic	69				18	question	96			
19	come	370				19	charge	84				19	application	98			
20	charge	84				20	time	85				20	judgment	99			

Corpus Analysis - collocates



modifiers of X="criminal"

	LogDice (strength of the collocation)
dangerous	10.86
hardened	12.14
unknown	11.35
hardcore	11.24
habitual	11.09
such	5.78
professional	10.3
feared	10.24
previous	8.08
bad	7.78
alleged	7.02
other	4.71

Corpus Analysis

Left context	KWIC	Right context
ough the court did not comply with the proviso to	section 251 of the Criminal Procedure and Evidence Code	, the accused persons admitted to the correctness of
the trial court did not have regard of the proviso to	section 251 of the Criminal Procedure and Evidence Code	before entering a plea of guilty, and I ordered a
nd charged with the offence of murder contrary to	Section 209 of the Penal Code	(Cap 7 : 01) of the Laws of Malawi .
ase they must prove the following elements under	Section 209 of the Penal Code	(Cap 7 : 01) of the Laws of Malawi ; (a) (b
as opposed to accidental or unlawful death under	Section 208 of the Penal Code	. Malice aforethought which is the m
is the mens rea for murder has been defined in	Section 212 of the Penal Code	as an intention to cause the death of or to do gri
appellant was convicted of Defilement contrary to	section 138 (1) of the Penal Code	and later sentenced to ten years IHL .
a charge of the offence of defilement contrary to	section 138 (1) of the Penal Code	he was found guilty and convicted . Or
a of guilty without having regard to the proviso to	section 251 of the Criminal Procedure and Evidence Code	. In support of this argument the appel
the statutory defence contained in the proviso to	section 138 of the Penal Code	before the plea taking process as well as the provis
the plea taking process as well as the proviso to	section 251 (2) of the Criminal Procedure and Evidence Code	. As was noted in the case of Marvellous Masa
lic Criminal Appeal No. 21 of 2017 the proviso to	section 251 (2) of the Criminal Procedure and Evidence Code	' must be respected religiously because it is pla

Law Citations - Annotations

- "They contended that between the time of conviction and sentence, **section 283 of the Penal Code** was amended, more particularly, in respect of the sentencing regime. Their argument, and therefore their case, was based on the change, or amendment to Section", [(65,94,'LAWCIT')]
- "The starting point, in our view, should be our General Interpretation Act . **Section 14 of the General Interpretation Act** states as follows: where a written law repeals and re-enacts with or without modification, any provisions of any other written law", [(77,121,'LAWCIT')]
- "**Section 14 of the General Interpretation Act** is a general saving section. would be very slow to speculate on the applicability of the doctrine of lex mitior in the absence of full arguments on", [(45,76,'LAWCIT')]

Case Citations - Annotations

CaseRef

Rep vs. Lungu [1995] 2 MLR 712 at 713

Republic vs. Sinos Pasipanadya Criminal Case No. 41 of 2008

Rep vs. Samson Matimati, Criminal Case No 18 of 2007

R. vs. Howeson cited in Pandirker vs. Republic

Towera Chitsa & Another vs. Republic, Misc. Criminal Application Case No. 160 of 2009

Republic vs. Bright Jamali, Confirmation Case No. 421 of 2013

R vs. Daveport (3) 88 cr App. R. at 41

Republic vs. Howeson cited in Pandirker vs. Republic was decided in 1936

Rep vs. Symon Kamuna Confirmation Case No. 669 of 2002

Fadweck Mvahe vs. Rep, MSCA Crim Apl. No 25 of 2005

Charles Khoviwa vs. Rep. Criminal Case No. 6 of 2007

Chiteta vs. R, 1960 R. & N. 199, at 205

We generated several hundreds of training examples containing case references.

ICCS Level 1

Table 1. Level I sections of the ICCS.

1	Acts leading to death or intending to cause death
2	Acts leading to harm or intending to cause harm to the person
3	Injurious acts of a sexual nature
4	Acts against property involving violence or threat against a person
5	Acts against property only
6	Acts involving controlled psycho-active substances or other drugs
7	Acts involving fraud, deception or corruption
8	Acts against public order, authority, and provisions of the state
9	Acts against public safety and state security
10	Acts against the natural environment
11	Other criminal acts not elsewhere classified

The International Classification of Crime for Statistical Purposes (ICCS) was developed by a task force of the UNODC and the United Nations Economic Commission. To our knowledge ours is a first effort in mapping the Malawi Penal Code to the ICCS.

Features for ICCS Level

- **Example:** (Level 1 Section 2) serious assault = ‘intentional or reckless application of serious physical force inflicted upon the body of a person resulting in serious bodily injury’, which, at minimum, includes gunshot or bullet wounds; knife or stab wounds; severed limbs; broken bones or teeth knocked out; internal injuries; being knocked unconscious; and other severe or critical injuries’ ‘serious physical force includes being shot; stabbed or cut; hit by an object; hit by a thrown object; poisoning and other applications of force with the potential to cause serious bodily injury’.

Manual Classification of 2019 Cases

NO	Case	ICCS Level 1	Type of Crime	Name of delict in text	Section in statute	full match section = ICC category	Rules of thumb
1	<i>MWHC/Kalemba</i> , The Republic v Humphrey Elia, Mary Nankhuku, Criminal Case No. 164 of 2018, 26 April 2019	01	intentional homicide	murder	Section 209 of the Penal Code	mostly	absence of euthanasia
15	<i>MWHC/Kamwambe</i> , Jamson Chagomeranda and Chikumbutos Madera v the Republic, Criminal Review no. 9 of 2018, Judgment 10 January 2019	02	assaults and threats	wounding	Section 235 (a) of the Penal Code	mostly	
16	<i>MWHC/Kamwambe</i> , The Republic vs Charles Mbengwa, Criminal Case No. 148 of 2018	02	assaults and threats		Section 235 (a) of the Penal Code	mostly	
17	<i>MWHC/Ntaba</i> , The Republic vs Maxwell Matchina Sosola and 11 others; Homocide case No. 13 of 2018;	02	trafficking in persons		Section 14 of the Trafficking in Persons Act	yes	
18	<i>MWHC/Kamwambe</i> , Nzeru, Dyson v The Republic; Criminal Appeal No. 32 of 2018, Appeal 14 March 2019	03	rape	defilement	Section 138 of the Penal Code (not in text)	yes	

Manual Mapping Penal Code to ICCS Level 1

Section	Statue / Law	Heading
Intentional Homicide	Section 208 of the Penal Code Section 211 of the Penal Code Section 213 of the Penal Code Section 215 of the Penal Code	Manslaughter
Intentional Homicide	Section 209 of the Penal Code Section 210 of the Penal Code Section 212 of the Penal Code Section 213 of the Penal Code Section 215 of the Penal Code	Murder
Intentional Homicide	Section 230 of the Penal Code	Infanticide
Attempted homicide	Section 223 of the Penal Code Section 400 of the Penal Code	Attempted murder
Attempted homicide	Section 224 of the Penal Code Section 400 of the Penal Code	Attempted murder by convict

MWCC Corpus

- Covers 665 judgments over 2010 – 2019 one folder for each year. Each judgment has associated a text file with the introduction part and an xml file in TEI format with metadata and paragraphs.

```
<?xml version="1.0" ?>
<TEI.2 id="judg_2010_1" lang="en" n="2010_1">
  <teiHeader date.created="2020-04-20" lang="en">
    <fileDesc>
      <titleStmt>
        <title type="full">
          <title type="main">Lawrence Chibwana Vs The State</title>
          <title type="sub">Criminal Appeal No 42 of 2010</title>
        </title>
      </titleStmt>
      <extent>113 paragraphs segments</extent>
      <publicationStmt>
        <publisher>Amelia Taylor, tNyasa Data Labs</publisher>
        <distributor>Amelia Taylor, tNyasa Data Labs, avtaylor14[at]gmail.com</distributor>
        <availability>COND AVAILABILITY</availability>
        <date>April, 2020</date>
      </publicationStmt>
      <notesStmt>
        <note>Only the court cases held at the Malawi Judiciary are deemed authentic.</note>
      </notesStmt>
      <sourceDesc>
        <bibl>Malawi Criminal Cases Corpus, covering 2010 - 2019, Copyright: Amelia Taylor.</bibl>
      </sourceDesc>
      <projectDesc>The corpus was part of a project "Semi-automatic Tools for Meta-Data e
received a grant from AI4D.</projectDesc>
      <classDecl>
        <taxonomy>
          <category xml:id="casetype">
            <catDesc>The type of case: order, bail, application, review, etc.</catDesc>
          </category>
          <category xml:id="courtofhearing">
            <catDesc>Criminal and which courts heard the case.</catDesc>
          </category>
        </taxonomy>
      </classDecl>
    </fileDesc>
    <textclass>
      <catRef target="#casetype">
        <keywords>
          <list type="casetype">
            <item>JUDGEMENT</item>
          </list>
        </keywords>
      </catRef>
      <catRef target="#courtofhearing">
```



MWCC Corpus

```
<catRef target="#courtofhearing">
  <keywords>
    <list type="courts">
      <item>REPUBLIC OF MALAWI</item>
      <item>MALAWI JUDICIARY</item>
      <item>IN THE HIGH COURT OF MALAWI</item>
      <item>MZUZU DISTRICT REGISTRY</item>
    </list>
  </keywords>
</catRef>
</textclass>
</teiHeader>
<text>
  <front>
    <list type="caseinfo">
      <item>CRIMINAL APPEAL NO 42 OF 2010</item>
      <item>CRIMINAL CASE NO 120 OF 2010</item>
      <item>BEING KARONGA FGM COURT</item>
    </list>
    <list type="parties">
      <item>LAWRENCE CHIBWANA</item>
      <item>THE STATE</item>
    </list>
    <list type="coram">
      <item>HON MR JUSTICE D T K MADISE</item>
      <item>Mr L B T Ndovi Counsel for the Appellant</item>
      <item>Mr Nicely Msowoya Counsel for the State</item>
      <item>Mr I Zimba Bondo Official Interpreter</item>
      <item>Mrs F Silavwe Court Reporter</item>
    </list>
  </front>
  <body>
    <p n="2"> INTRODUCTION. The Appellant in this matter Lawrence Chibwana was convicted of the offence of bringing in property dishonestly acquired outside the country contrary to section 331, Penal Code. It was alleged that the Appellant on or about the 24th day of June, 2009 at Songwe Border post in Karonga brought into Malawi a Scania truck motor vehicle with chassis No 4594569 and Engine NO 1 387254, 62 Fridges and 100 compressors, all valued at K2,491 ,000.00.</p>
    <p n="3">The allegation was that the said properties were dishonestly obtained by the Appellant in Dar-es-Salaam, Tanzania by transferring the title of the said properties from the name of the consignee a Mr. Karhakubwa Dieu, Donne of Democratic Republic of Congo into his own name after paying shipping charges, thereby evading the purchase price.</p>
    <p n="4">The Appellant denied the charge and after a full trial he was found guilty, convicted and sentenced to a prison term of 12 months but suspended for 24 months. Being unsatisfied with the decision of the lower court, he now appeals to this Court against conviction, sentence and order.</p>
    <p n="5">GROUNDS OF APPEAL The Appellant filled 19 grounds of Appeal which we reproduce as</p>
    <p n="6"> a) The Learned Magistrate erred in law in convicting the accused against the weight of evidence. b) The Learned Magistrate misdirected himself in rejecting on the wrong premise in law the admissibility of the London Interpol Report which exculpated the Appellant and could have lead to his acquittal. c) The Learned Magistrate misdirected himself and in the process erred in law in concluding that the burden of proof had been discharged. d) That the Learned Magistrate, with respect to him grossly misdirected himself on the standard of proof expected by the court from the prosecution. e) The Learned Magistrate erred in law in starting from the presumption of guilt and ignoring the constitutional
```

Work in Progress

- **Where we are:**

- We created the MWCC corpus
- We have generated hundreds of annotations to be used for training
- We mapped the Penal Code to ICCS Level 1
- We classified the 2019 cases on ICCS Level 1 – training set

- We are still working on key legal terms for the ICCS classification

- We are working on creating word2vec models for the corpus

- **Challenges & Opportunities:**

- It is not trivial to select legal keywords for the ICCS classification
- TF/IDF may not be a sufficient measure of Significance
- We will use a combination of keywords and Citations
- Hence the importance of annotations and mapping to the ICCS
- Challenges to do with the structure of the documents, e.g., how to treat footnotes (many contain case / law citations)
- Documents are large: how to chunk them for classification?

Thank you.

- We thank AI4A for the support of this project.
- We thank you for your attention.

Contact: ataylor@poly.ac.mw / avtaylor14@gmail.com

- Questions?